Industrial Data and Regional Economic Development
Phase 1 Study and Findings

Dr. Jennifer Clark, Georgia Institute of Technology

Dr. Thomas Lodato, Georgia Institute of Technology

Supraja Sudharsan, Georgia Institute of Technology

February 2018

Georgia Tech Energy Policy and Innovation Center

About the Energy Policy and Innovation Center

The Energy Policy and Innovation Center (EPICenter) was launched in the Fall of 2016 with the mission of conducting technical research, providing information on various contemporary topics in the energy field, and coordinating activities among leaders and innovators across industries and sectors. The Center explores the intersection of policy and technology, while leveraging the extensive expertise present across firms, research institutions, policymakers, and other government and non-government organizations in the Southeastern United States.

The Center is the first known implementation of a regional partnership to focus on the interdependencies of energy policy and technology in developing and implementing significant, cost-effective, and market-based carbon reductions. An assortment of deliverables will be produced by the Center, including but not limited to work products, events, educational outreach, and workforce development. Through these outputs, the Center strives to help accelerate a variety of reliable, affordable, and low-carbon energy options in the Southeast.

In executing its mission, EPICenter draws upon voluntary contributions from external organizations. The center is funded by an endowment and annual cash gifts to the Georgia Tech Foundation, and receives additional support in the form of personnel time and other in-kind contributions. Input from external entities that accompanies support, including recommendations related to center studies or operations, is subject to the discretion of EPICenter leadership. Similarly, no particular work product, findings, or implied results of center deliverables shall be linked, or give the perception of being linked, to a specific donation by any individual participant.

Richard Simmons, PhD.
Director of EPICenter

About the Center for Urban Innovation

Georgia Tech's Center for Urban Innovation (CUI) combines technology and policy through interdisciplinary research on smart, sustainable cities and regions. CUI conducts interdisciplinary, scholarly research on urban and regional economic and policy issues with an emphasis on the expanding role of institutions in designing and governing resilient regional economies. Uniquely positioned within a world-leading technological research university and in a global city, CUI was launched in 2014 to pursue research-driven discoveries that engage challenges and economic opportunities faced by cities and regions.

Jennifer Clark, PhD.
Director of CUI

Preface

The digitalization of industries is leading to new services and opportunities for firms, regions and economies. The ability to monitor and diagnose equipment conditions and operations in real-time, respond to changes in equipment performance, and share this data across stakeholders and regulators is expected to improve business profitability, innovation, job growth, and regulatory compliance. The goal of this study is to evaluate the nature of these changes as they relate to the production and use of industrial data. Simultaneously, the study identifies and assesses the factors that influence regional economic development.

The research presented here is a collaboration between the Center for Urban Innovation (CUI) and the Energy Policy and Innovation Center (EPICenter) at the Georgia Institute of Technology. This partnership builds upon CUI's expertise on data and regional economic development and the EPICenter's access to the energy industry and experience in the field. The goal of this research is to delineate the evolution of the industrial data industry and its implications for regional economic development in the Southeastern United States. The research uses an industry studies methodological approach, building empirical cases of the use of industrial data in firms across sectors. Through key informant interviews and extensive primary research, this analysis builds a production circuit of the industrial data industry and draws out those factors that explain the location of these processes globally. By considering the geography of processes associated with a worldwide industrial data supply chain, this research contributes to the literature on globalization and provides implications for regional economic development policy.

Jennifer Clark, Director of the Center for Urban Innovation, and Tim Lieuwen, Executive Director of the Energy Policy and Innovation Center lead this effort, supported by Richard Simmons, Director of EPICenter, Thomas Lodato, Research Scientist at CUI, Supraja Sudharsan, Graduate Research Assistant at CUI, and Kerri Metz, Research Associate at EPICenter. We thank our EPICenter partners for their time and assistance, as well as their availability and feedback while conducting the study, and look forward to their continued support. We thank our Georgia Tech colleagues – Dr. Nagi Gebraeel, Professor Peter Swire, Dr. Annie Anton, and Jesse Woo (attorney and privacy research associate), who provided insight into data analytics, and related governance and legal issues around industrial data.
This report contributes to the EPICenter's objective of supporting interdisciplinary research that stimulates innovation in energy policy and technology throughout the Southeast region. EPICenter intends to publish the findings of the study and share the report with the Center's sponsors and participants.

Table of Contents

INDUSTRIAL DATA: WHAT ARE THEY AND WHY DO THEY MATTER? 5

BACKGROUND 5

Introduction to Study 6

RESEARCH APPROACH AND METHOD 7

Selection of Case Studies 9

ANALYSIS: ID PRODUCTION IN POWER GENERATION 10

Sources and Uses of ID 10

Geography of ID Production in Power Generation 11
- Firm development 11
- Reliable Connectivity 12
- Domain Knowledge, Data Scientists, and Research Institutions 12
- Data Governance 13

FINDINGS: ARTICULATION OF AN EMERGING INDUSTRY 15

ID Production Circuit 16

Regional Competitive Advantage Drivers 16

Emerging Industrial Data Ecosystem 17

CONCLUSION: PRELIMINARY STRATEGIES AND RESEARCH AGENDA 17

Recommendations 19

BIBLIOGRAPHY 19

Appendix A 21

.... 23

Appendix B 23

Industrial Data: What Are They and Why Do They Matter?

Industrial data (ID) are data obtained by measuring and assessing the production and operations of industrial equipment, processes, and systems. These include:

1. **Production data:** Production data are data from industrial equipment, processes, and systems *during production*. For example, production data includes data from advanced manufacturing industries produced *during* manufacturing by factory equipment, logistics networks, and human-resource systems.

2. **Operations data:** Operations data are data from industrial equipment, processes, and systems *during operation*. For example, operations data from the energy sector may include data from power generation assets (e.g. turbines) in operating power plants, transmission and (non-consumer) distribution networks (e.g. substations and transmission lines), and service and maintenance records, amongst many other sources.[1]

The study is meaningful for several reasons:

1. ID are predicted to be integral to a fourth industrial revolution (Schwab, 2015). The coalescence of operational, information, and computation technologies are garnering a great deal of attention among industry, academic and policy communities. Recent studies on the subject posit that by enabling data-driven decision-making, this fourth industrial revolution will increase productivity, innovation, and recast opportunities across economies and society (Bernstein and Raman, 2015; World Economic Forum, 2016b).

2. By transferring knowledge embedded in the workforce into machine learning models,[2] the emerging ID industry is expected to lead to a highly-skilled workforce, predictive approaches to decision-making, and an economy and society organized around these forces (Kitchin, 2014a; The Economist, 2010; World Economic Forum, 2016b).[3]

3. Stemming from these expectations, a number of government, industry, and hybrid initiatives have emerged in recent years, namely, "Industrie 4.0",[4] Industrial Internet, Advanced Manufacturing, and Smart Manufacturing.

By delineating the production process and clustering characteristics of ID, this study considers strategies for regional economic development in the Southeastern United States. In addition, it proposes a research agenda to explore the emergence of ID and its implications for industries, firms, and institutions.

Background

Data are ubiquitous. Digitalization of production and operations has transformed equipment, processes, and their interactions into data-generating objects and systems that extend across physical and digital spaces (World Economic Forum, 2016a). Sensors and actuators embedded in equipment connect to each other, and connect equipment to common interfaces, databases, and services via wired and wireless networks. Through wired and wireless communication

[1] Not considered here are smart grid uses that involve collection of data on consumer behavior, or data from consumer devices. That is, ID differentiates proprietary data generated by industrial equipment, from private, consumer data.
[2] Han et al. (2012); Hastie et al. (2009).
[3] Data-related services such as monitoring equipment's operation, and predicting its performance in advance, are expected to improve productivity, transform business models, redraw geographies where value is captured, and give rise to an economy with a significant dependence on data-related firms and processes.
[4] "Industrie 4.0" is an initiative sponsored by the German government to support advanced manufacturing through the development of cyber-physical systems.

protocols, as well as data architectures, platforms, and services, a high degree of interoperability between the physical and digital space is possible. As such, data can be collected, distributed, compared, and aggregated, and so on, leading to the generation of even more data.

However, the proliferation of data itself is not a novelty. Large volumes of data have existed in consumer and financial service industries long before the rise of industrial data (Kitchin, 2014a). For instance, customer transaction data from Walmart reached 2.5 petabytes, according to an Economist report in 2010 (The Economist, 2010). Facebook, unsurprisingly, processes billions of data points in content and photos daily (Constine, 2012). Indeed, by all accounts, the size of ID is smaller than that of consumer data (O'Donovan et al., 2015).

What is new, however, is the development of data infrastructure for the collection, transmission, and storage of data, as well as the systems, algorithms, and services that have allowed data to be understood through aggregation, analysis, and visualization. Machine learning techniques have expanded the ability of firms to handle complex data, aggregated from an array of sources (Kitchin, 2014b).[5] By mining data for the purpose of detecting patterns and predicting outcomes for various use cases in the industrial domain (Han et al., 2012; Hastie et al., 2009), these techniques have arguably led to a "new dimension of collecting raw data over a distance and its aggregation/analysis, where [the] value of data increases with [its] aggregation level" (Konig, 2016). The development of cloud computing, embedded computing, software systems, and cyber security technologies are all shaping the development of the emerging ID industry (Möller, 2016; Wee et al., 2015).

Introduction to Study

The study has to date focused on understanding ID arising from electric power as an exemplary industry sector. Within this scope, ID from power generation assets, such as power plants, wind, and solar generation assets, etc., are considered. By collecting industrial data use cases, this study defines the scope of industrial data, its sources, and its uses in power generation. This scope provides an empirical basis for understanding the origins and evolution of industrial data as a product and as an emerging industry. Additionally, the scoping informs firms and policymakers about prevailing sectoral linkages of ID. Three goals guide this study:

1) Demonstrate the emergence of a new industry;
2) Illustrate the geographical underpinnings of firm strategies in ID; and
3) Unpack linkages at the power generation-data nexus.

This report is divided into four sections. The first section explicates the research approach undertaken for this study. This section also introduces the analytical framework. The study considers firm strategies as events of analysis and traces their evolution over space and time.[6] It explains the variation in strategies among firms across space and time, across large and small firms, and across firms operating within different institutional structures. Sections 2 and 3 delineate the geography of ID production, and its sectoral linkages. Finally, it concludes by discussing the characteristics of the

[5] In general, big data refers to "large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future" (Executive Office of the President, 2014). Rob Kitchin (2014) provides a comprehensive definition of big data as "huge in volume, consisting of terabytes or petabytes of data; high in velocity, being created in or near real-time; diverse in variety, being structured and unstructured in nature; exhaustive in scope, striving to capture entire populations or systems (n = all); fine-grained in resolution and uniquely indexical in identification; relational in nature, containing common fields that enable the conjoining of different data sets; flexible, holding the traits of extensionality (can add new fields easily) and scale-ability (can expand in size rapidly) (see Boyd and Crawford, 2012; Dodge and Kitchin, 2005; Laney, 2001; Marz and Warren, 2012; Mayer-Schonberger and Cukier, 2013; Zikopoulos et al., 2012)."

[6] Firm strategy refers to the organizational routines, processes, and decisions made by firms in response to changing local, regional and global conditions. This study considers strategies that involve both operational effectiveness, which Porter (1996) contends is a short-term result that is easily imitated by rivals, as well as activities undertaken by firms that are unique. This study accounts for the dynamic nature of firm strategies. That is, firm strategies change over time, due to interaction with other local, regional, and global conditions. For more details, see the discussion of literature on evolutionary economic geography.

emerging industry, and major questions to be tackled in order to understand its regional clustering characteristics for economic development in the Southeastern United States.[7]

Research Approach and Method

By adopting a generalized definition of ID—namely, ID as data obtained by measuring and assessing the performance of production and operations of industrial equipment—this study considers both the sources and uses of data in the emerging ID industry. In so doing, it uses Dicken, (2007)'s global production network and Porter's diamond model (Porter, 1990, 1998a) as analytical frameworks. The global production network allows us to explore the interactions between firms, institutions, and other actors at different levels of the global economy (Coe et al., 2008). In conjunction with Porter's diamond model,[8] it situates the study in the transnational space, without compromising on territorially specific conditions. See Figures 1 and 2 for further explanation.

Through this approach, we map firm-specific use cases of ID in power generation, while considering constraints at the level of the industry, region, and nation-state. The research method for this study is presented in Table 1 and conceptually explained in Figure 3. Corporate publications, industry trade magazines, and literature from operations research and industrial engineering were all utilized to develop a model of the emerging ID industry. Studies and key informant interviews with large, incumbent industrial firms, utilities, SMEs, start-ups and non-profit organizations were used to build a case study of ID production and factors that influence ID production in power generation.

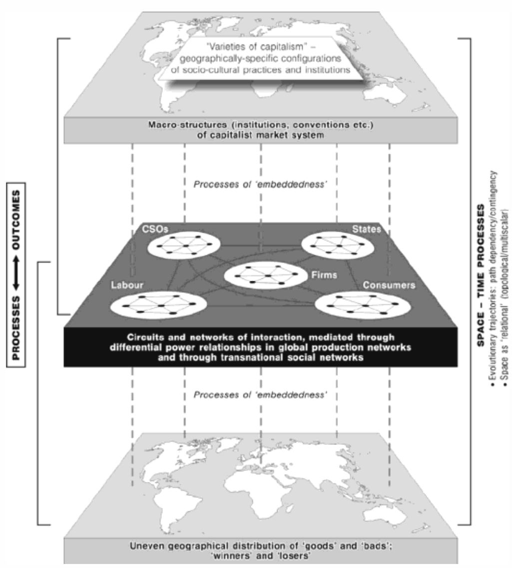


Figure 1. Conceptual map of global production network. This framework captures both the territorially-embedded processes, and the macro-structures in the international system that influence the outcome of economic processes globally. Source: Dicken (2007).

[7] This study considers the states of Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee to be in the "Southeast." These are defined based on a combination of megaregions and common utility territories present in the region.

[8] Porter (1998b) developed the concept of "clusters" which refers to linkages between firms, supporting organizations, institutions, infrastructure, and policymakers within a certain region. A widely-known example is the technology cluster in Silicon Valley. Clusters are used to provide a way of focusing economic policy in a region. A cluster approach to regional economic development is a common framework of study, although widely debated in certain instances (for instance, whether it is fruitful to build a regional cluster via policy intervention). However, the existence of relevant firms, supporting infrastructure, and organizations in the Southeast renders this a useful approach for the purposes of this study.

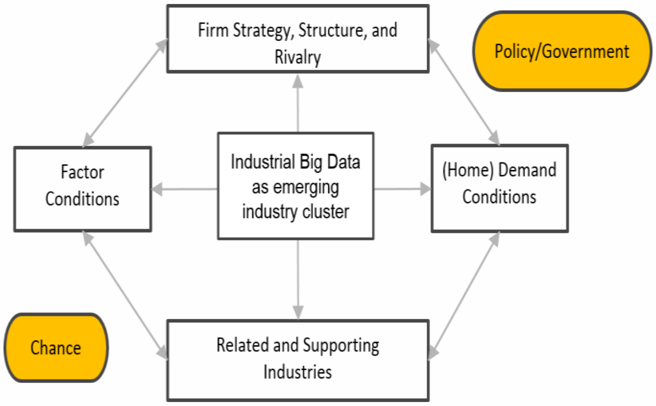


Figure 2. Conceptual map of Porter's diamond model for competitive advantage of nations. Adapted from Porter (1990, 1998a)

Table 1. Research approach undertaken by team for industrial data study

Research Component	Source	Outcome
Literature survey to understand sources and uses of industrial data	Corporate publications, industry trade magazines, literature from operations research, industrial engineering, systems engineering, think tank reports.	Spreadsheet of use cases based on firms operating in power generation.
Key informant interviews	Large, incumbent producer firms, SME, start-up, non-profit organization operating in power generation.	Case studies of firm strategies in power generation-data nexus
Preliminary stakeholder review, and phase 1 in-team review	Roundtable presentation and discussion with EPICenter's industry stakeholders, review of preliminary report.	Crystallize energy-data linkages in study.

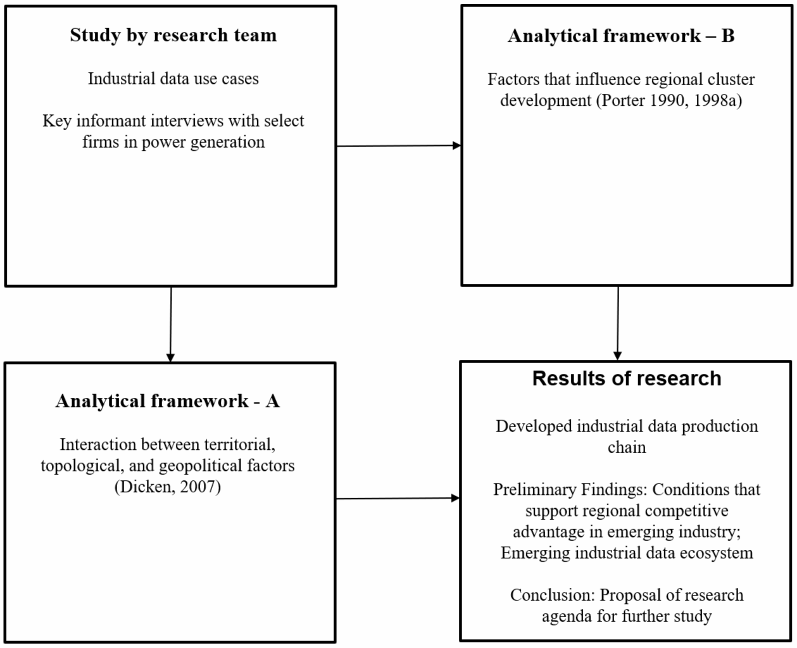


Figure 3. Conceptual map of research approach. Developed by research team.

Selection of Case Studies

As the events of analysis are firm-level strategies, we consider firms across the power generation value chain, namely, equipment (ex: turbine) suppliers, power suppliers, independent power producers, vertically-integrated utilities, energy management firms, automation firms, and more. These are listed in Table 2. However, this list not exhaustive. Traders, power plant commissioning firms, etc., are not considered for the study.

Table 2. List of Power Generation Firms (by Type) and Inclusion/Exclusion Criteria for Study

Firm type	Included in study?	Explanation
Electricity Supplier	Yes	1.ID use cases from independent power producers (including sectoral use cases in wind, nuclear, power generation). 2.ID use cases in vertically-integrated utility.
Equipment provider	Yes	ID use cases and case studies from tier-1 equipment suppliers.
User – Industrial	No	ID case study from industrial automation provider as proxy.
Trader	No	This is beyond the scope of the current phase of the study. The current phase only includes firms that are directly associated with power generation.
ISO	No	Considered as part of the institutional landscape, within which ID production in power generation is emerging.
Regulator	No	Considered as part of the institutional landscape, within which ID production in power generation is emerging.
Services provider – energy management	Yes	1.ID use case from firm that provides energy management software for industrial customers, data centers. 2.Case study of non-profit group representing energy management companies as proxy.
Services provider – tech	Yes	ID case studies from data aggregator, platform developer, software and managed services provider.

Analysis: ID Production in Power Generation

ID production includes both the sources and uses of ID. For instance, in power generation, data are generated from assets such as gas turbines, wind turbines, and solar panels as they are being produced, and following installation. Other sources include data from controllers, manufacturing execution systems, and other plant sources, as well as distributed generation (including renewables deployed at an industrial customer's site (e.g. "behind-the-meter" data)).[9]

Sources and Uses of ID

From the selected firms (see Table 1 & 2), the research team developed a typology for classifying firm strategies in terms of their sources and uses of data. Table 3 shows some examples of use cases undertaken by firms operating in power generation. It reveals that there are four types main types of uses associated with ID. Firms use data to:

1) Monitor equipment performance real time or otherwise,
2) Diagnose equipment performance by comparing with historical and baseline conditions,
3) Optimize operations, and
4) Develop new markets.

In response to changes, and in response to a cost-benefit analysis, they also optimize equipment performance (inductive or predictive maintenance), In some cases, this leads to the development of new markets for existing firms, as well as spawning new firms in the ID ecosystem. [10]

Table 3. Examples of firm-level use cases in industrial data production.[11,12]

Firm Name	Asset performance management			New Services/ Markets
	Remote Monitoring	Diagnostics	Optimization	
Utility	X		X	
General Electric – Power	X	X	X	X
Vestas – Renewable energy	X	X		
Nextera – Distribution	X	X		
Deutsche WindTechnik – Renewable energy	X	X		
Trove Predictive Science – Tech	X	X	X	
Urjanet (Data integration for)	X	X	X	X

Figure 4. depicts the collection and use of data in power generation. As seen in the figure, internal and external data collected from power generation assets and the surrounding environment, respectively. Strategic partnerships enable firms to store and aggregate data, which are then analyzed and used for monitoring, and visualizing present and future equipment performance.

[9] Based on key informant interviews.
[10] For example, UrjaNet integrates industrial customers' utility data, and sells insight from this data for credit scoring, and sells utility billing data to lenders. See Shieber (2017).
[11] The table is based on firm strategies as events of analysis. The table includes type and nature of firm operation in the specific industry. Other columns (not displayed here) include the sources of data collection and strategic partnerships undertaken by firms to pursue industrial data uses. Developing a spreadsheet of sources and uses enabled the team to map the variation across the processes and outcomes associated with the emerging ID industry.
[12] Developed by research team. Also see General Electric (2017), Boria (2015), Plataine (2017).

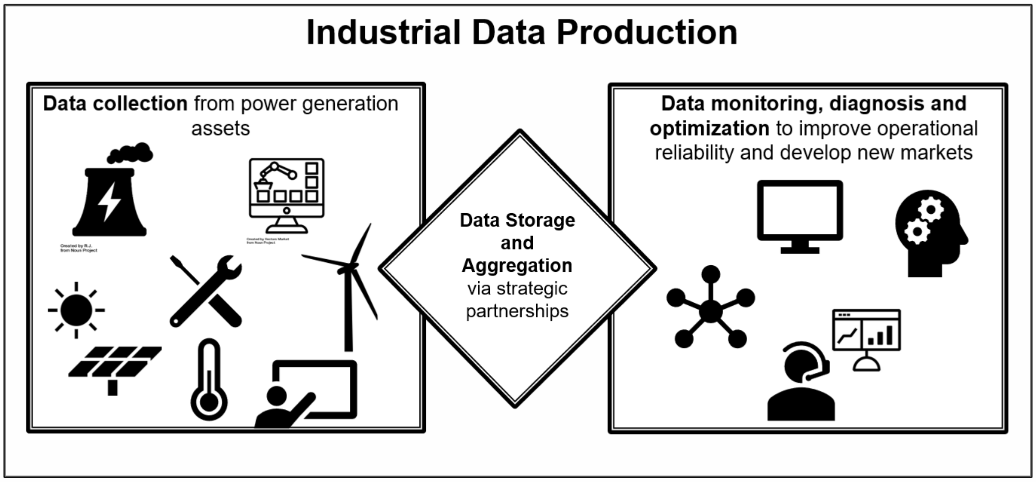


Figure 4. Industrial Data Production Process in Power Generation.[13]
Image credits: Vectors Market and R.J. from Noun Project.

Geography of ID Production in Power Generation

The previous section situated the study in terms of the sources and uses of ID. This section addresses the question – under what conditions do firms evolve into the ID ecosystem? The analysis presented here is based on key informant interviews with a few representative firms in power generation. See Appendix A to understand the conditions that influence the evolving ID ecosystem.

Firm development

Firms evolve in different ways across the ID production process. Some incumbent, industrial firms such as GE, Siemens, and others operate across the ID production process. Others, such as Southern Company, operate across select ID production activities. Still others, namely SMEs and start-ups develop capabilities to provide specific services such as data aggregation and analytics for utilities and other energy users. These firms also utilize ID to access new markets, and subsequently, to expand their businesses and service offerings.

This variation in firm development is attributed to firm strategies, and the evolution of these strategies over time. For instance, as tier-1 equipment providers in advanced countries face decreasing revenue opportunities, they harness the developments in data-driven technologies to pursue new profit opportunities and secure a competitive advantage (Tweed, 2010). Utilities undertake monitoring of power plant operations to ensure reliability of operations, due to the proprietary nature of data, and in some instances, due to perceived costs and benefits associated with firm-level technical capabilities.

However, the nature and sophistication of these operations are determined by internal firm capacities. Firms rely on a number of strategic partnerships with tech and automation firms, and maintenance service providers, to pursue ID use cases. For instance, Southern Company monitors its fleet operations in Birmingham, Alabama, via partnership with Schneider Electric (Bray, 2015). Duke Energy monitors its renewable power operations from Charlotte, North Carolina (Duke Energy, n.d.). NextEra has located its operations in Orlando, Florida, to cut costs stemming from their a dependence on the manufacturer for maintenance (EnergyNow, 2015). Lastly, Duke Energy monitors its own wind operations along with DTE Energy's renewable operations across the United States. Here again, the nature of operations is determined by internal firm strategy.

[13] Developed by research team based on key informant interviews. Data are collected from generation and other power plant assets (left). These are aggregated with weather data and data from existing SCADA systems, to develop new insights about equipment performance (right). The production processes are enabled by strategic partnerships between tech providers, equipment providers, utilities and power producers for sharing data, developing analytical techniques, integrating data across operation systems, etc.

In Appendix A, the location of the monitoring and diagnostics operations of large, incumbent, industrial firms A, B, and C were determined by firm evolution over time. Other factors have also supported their location over the years, including reliable connectivity, domain knowledge, and data governance. These factors are discussed in detail below. The factors not only explain the geographic evolution of large, incumbent, industrial firms but also explain the evolution of tech firms, SMEs, start-ups, and the ecosystem of supporting institutions around ID.

Reliable Connectivity

Connectivity determines the cost of communication among processes and personnel within and across industrial environment and regions. Within our research, we define connectivity by the speed of interaction between machines, and between machines and personnel enabled by the installation of information and communication technologies. Therefore, connectivity, in this sense, is determined by the nature of the communication protocol that is in use, which may be based on wired or wireless devices, provided via internet (cable, fiber, or satellite) or Bluetooth or any other LAN or low-power WAN (Elgendy and Elragal, 2014).

Connectivity and the reliability of connectivity are intrinsically related to the economic geography of ID. On the one hand, infrastructure in global cities (Sassen, 2001), allow for greater mobility of information and communication in these regions. On the other hand, rural and agricultural regions, even in industrialized economies such as the United States, experience poor broadband connections (Tomer et al., 2017).

Illustrated through its relationship with geography, connectivity influences how ID uses evolve across different industries. In resource-intensive industries, such as oil and gas, where refining operations are built either close to the source or close to markets, the reliability of connectivity determines where ID production processes are located (Marr, 2015; University of Cambridge, 2016). Even in data-intensive industry sectors, connectivity is likely to influence geography of data collection, storage, aggregation, and use, and vice-versa. This is evident in wind and solar industry sectors, where location of large-scale turbines and panels are determined by availability of resources, proximity to the grid, and other land-use regulations (Rodman and Meentemeyer, 2006). Therefore, the interaction between connectivity and firm strategies in the emerging ID industry is a function of existing sectoral geographies. See *B1: Connectivity and Firm Operation in Algeria*[14] below for an example.

B1: Connectivity and Firm Operation in Developing Regions of the World
Reliable connectivity is a challenge in a number of resource-rich, developing regions of the world. . Reliability of connectivity influences the time taken between data collection and their analysis. Where there is a delay between data collection and use, interviews revealed that local personnel are used extensively to support data flow and to communicate the results of diagnostics.

Domain Knowledge, Data Scientists, and Research Institutions

The interviews further reveal that domain knowledge is crucial to the evolution of the ID industry. Firms rely on different ways of acquiring knowledge across industry, and technology sectors. These insights are, in turn, utilized to develop algorithms and analyze ID. One approach adopted by firms to facilitate the transfer of this knowledge in the emerging ID industry is pairing domain experts with data scientists to develop algorithms and analytics models that address operational issues within the context of the industry.[15]

[14] Based on key informant interview.
[15] Key informant interviews with executive from large, incumbent industrial firms.

Research institutions and networks[16] serve as intermediaries in this process. Universities train students to be data scientists, imparting skills across computer science, statistics and engineering, the combination of which is materializing as a key component to support and build the emerging ID industry.[17] Through their partnerships, universities also provide exposure for students to industry-specific practices and strategies.

Other ways of acquiring domain knowledge include firm acquisition, collaboration with firms that have domain expertise, and investment in internal capacity development.[18] One instance of internal capacity development is skilling college students to develop a pipeline of talent in this industry.[19] Importantly, although establishing a pipeline illustrates the third strategy, it does not preclude this or other firms from using other strategies as well.

Therefore, domain-specific knowledge continues to be important, despite the rise in machine learning techniques.[20] Strategies undertaken by firms, institutions, and networks to transfer domain knowledge matter for the evolution of the emerging ID industry.[21] Moreover, research institutions serve as both providers of knowledge and incubating spaces for large and small firms in the emerging ID industry.[22] For an example, See *B2: Firm Location and Regional Academic Institutions* below.

Where firms possess internal capacity, existing firm geographies continue to be important. However, to develop new ID uses, as well as for new firms to enter these markets, location is determined by access to research institutions and associated incubating spaces.

B2: Firm Location and Regional Academic Institutions

According to a representative of a large, incumbent, industrial firm, their location in the Southeast was chosen for several distinctive reasons. Firstly, the educational excellence and density of organizations in data analytics operations in the region provides access to talent and help to develop strong ties with the innovation community. Secondly, Atlanta's airport hub serves the firm to locate the region as an epicenter for thought leadership in industrial data. Thirdly, the location of universities and firms at close proximity, as evident in the case of Georgia Tech and its nearness to the Technology Square is ideal for finding talent, developing ideas, and incubating companies, and collaborating with partners from the technology community.

Data Governance

Data Governance includes governing the quality and format of available data, the uncertainty associated with analysis of greater volumes of data,[23] establishing processes for ensuring transparency and accountability in data access and use, and dealing with proprietary, legal, and regulatory issues in the above processes (Malik, 2013).

These challenges exist at the firm level, but they vary by industry and by regional institutional structures. Addressing these issues has the potential to improve efficiency for firms, reduce the costs of regulatory compliance, and ultimately spur innovation.[24] Governance challenges also exist where legacy systems need to be integrated into modern

[16] Networks such as Industrial Internet Consortium and Smart Manufacturing Leadership Coalition enable collaboration between firms across the industrial data supply chain.
[17] Key informant interview with machine learning expert.
[18] Based on interview with SME (tech) firm.
[19] Based on interview with a SME (tech) firm.
[20] "Confidence" in algorithms goes down as size of data increases (interview with analytics expert from Georgia Tech). Therefore, building the analytics piece in the ID supply chain requires the input of domain experts to validate results from computing. Domain knowledge is also important in other ways. Interviews indicate that not all collected data are utilized for analysis and decision-making. Typically, all available data are collected in a data lake architecture. Personnel with domain knowledge expertise segregate useful data from others for further aggregation and analysis.
[21] Interview with expert.
[22] Interview with executive from a start-up firm.
[23] As the amount of sensor data increases, the uncertainty associated with predictive models also increases (per interview with analytics expert at Georgia Tech). Uncertainty is also associated with the complexity of data that are being analyzed. See (Malik, 2013) for more on this. See Hashem et al. (2015) and Loshin (2013) for more on the subject of data governance.
[24] For instance, in nuclear power generation, industry regulations that require frequent maintenance of pumps in nuclear power plants result in equipment that are not operational for significant periods of time every year.[24]

computing infrastructure. This has regional implications, depending upon the location of brownfield and greenfield installations (O'Donovan et al., 2015).

In power generation, contractual agreements between utilities or power producers and equipment suppliers determine how ID are shared and used. By collecting and aggregating data across an extensive global fleet of equipment, certain OEMs have an edge in shaping the types of agreements and partnerships that emerge within ID. Owing to their access to historical data, equipment know-how, some incumbent industrial firms are able to develop their internal capacity and evolve as front-runners in this emerging industry.[25] For new firms and ID uses to develop into this existing network of firm relationships, governance of data use, sharing, and access need to be resolved.

Governance challenges also exist where a public good argument is made for sharing data. It is argued that greater access would allow energy management firms that are not currently a part of regional ID production chain to innovate and bring more energy efficient products into the market (TechNet et al., 2016). This reveals the tension that exists in the issue of data governance, between utilities which may seek to retain their revenue, and energy management firms which seek to develop energy efficient products for their consumers. Where legacy systems are involved, there are also questions of who assumes the cost of investment to be made in computing infrastructure and its maintenance (Ferris and Rahim, 2016).[26]

In addition to firm-level contractual agreements and regional institutional structures, evolution of data access and ownership also depends upon the evolution of national institutional structures. Historical evolution of financial institutions in the United States and Europe have instilled differing risk aversion (Brecht, 2015). Therefore, regional institutional structures and policies may play a bigger role in determining who wins in the above-discussed scenarios, and thereby influence the trajectory of regional economic development.

In sum, data governance challenges are defined by spatiality of firm operations, structure of the industry in which the firm operates, and societal value associated with data sharing. Here, questions go beyond simply sharing data to which data need to be shared, with whom, and at what granularity and pricing across stakeholders. The proprietary nature of firm data dictates that firms and policy makers make decisions about who should *not* have access to ID as much as who should. These challenges only intensify as certain firms (ex: critical infrastructure such as hospitals, data centers), cities, and regions seek energy independence to ensure reliability of power services and strive to achieve other goals such as increasing energy efficiency[27] and using 100% clean energy.

B3: Grid Data Access

An executive from a non-profit organization advocating for data access rights explained the challenge of data access and sharing as follows (edited for clarity)

"Starting with the grid, we're in the midst of a bit of industry upheaval, and that is at the distribution level. For hundreds of years it has been men behind a current with millions of miles of voltage and wires that they've operated with little oversight and outside understanding, so having access to the assumptions they're making about the [grid]system and the models they use for that decision system promises to be a fundamental change. Who owns the data, why they are obligated to share it, and what rights do the different stakeholders have, are all important to consider in this space."

[25] Based on key informant interviews..

[26] Similar consideration goes into determining the granularity at which utilities need to share data. See Kusiak (2016).

[27] Based on interview with executive from large, incumbent producer.

National Data Policies

So far, we have seen that data collection, aggregation, storage, and use are undertaken globally, depending upon several conditions. Industry and equipment domain knowledge are aggregated in certain advanced economies. The evolution of firms, industries, regional institutional structures, and skilled labor that are required to transform this domain knowledge into actionable insights are disaggregated across different locations globally. Consequently, national data policies matter for the evolution of the emerging ID industry.

Data localization refers to the storage of data in the country from which it originated. Evolving data localization policies across nation-states influence where data are stored. Data from some power plants are transferred to local storage units before being made available for access via cloud due to the originating country's policies.[28] In 2015, South Korea enacted the Act on Promotion of Cloud Computing and Protection of Users, which includes rules of data localization for cloud computing networks serving public agencies (Cory, 2017), enabling diversity in firms' approaches to data collection in these regions. One instance of this is the use of local virtual private networks (VPNs).[29]

With the rise of cloud computing, national data policies also affect the ability of firms to take advantage of low-cost computing infrastructure. For instance, data localization policies in China restrict data center locations to the region, ignoring other low-cost options. One instance of the influence of data localization policies is that of Microsoft, which partners with a Chinese firm to extend its cloud services business in China (Verge, 2015). Cory (2017) estimates the cost of such data localization policies to be 0.1-0.36% of United States' GDP. Where there are no restrictions from national policies, the dynamics between the physical location of "clouds" and the legal ownership of data also raise questions about how disputes will be settled and who should possess legal jurisdiction over this data (Dlodlo, 2011).

Therefore, national data policies are critical in the evolution of the emerging ID industry and influence the location of data processes and associated investment, and in turn determine the winners and losers in the emerging ID economy. For other examples of countries' rationale for data localization, see *B4: Data localization under Cyber-Security Rationale* below.

B4: Data localization under Cyber-Security Rationale

Cyber-security and data-localization policies have a complicated relationship.

The Industrial Internet Consortium (n.d.) reports of examples of hacking in recent times – which indicate that denial-of-service, phishing and password theft are common forms of cyber-attacks. This leads to the conclusion that security threats exist at plant-level infrastructure, data centers, and cloud computing infrastructure.

Security concerns also arise from the nature of technology used in plant installations. For instance, digital instrumentation in nuclear plants are subjected to stringent regulations, owing to concerns over bugs in associated software, concerns over protecting data via robust encryption and authentication mechanisms, with wireless technology considered as the least secure among these (IAEA, 2011).

Findings: Articulation of an Emerging Industry

The above discussion reveals that the evolution of the ID industry depends upon the dynamics of firm-level, regional, and global factors. How firms respond to global changes, to competition at the regional level, and through organizational routines and processes are all influential in their evolution in the emerging ID industry. This section visually presents three findings from this study, namely, the production circuit in ID, drivers of regional competitive advantage in this emerging

[28] Interview with executives from large, incumbent, industrial firm.
[29] Key informant interview

industry, and the evolving ecosystem around ID. All of the above are developed in the context of the sources, uses and geography of ID in power generation.

ID Production Circuit

In Figure 5, data are collected from several internal and external sources. These are stored in local servers, or in the cloud. The collected data are then aggregated depending on intended use, analyzed, and finally used to understand equipment performance and generate value.

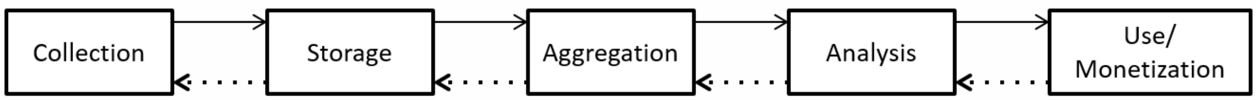


Figure 5. Industrial data production circuit.[30]

Regional Competitive Advantage Drivers

Four factors play a crucial role in influencing where firms locate their ID production activities, including collection, storage, aggregation, analysis, and use. These conditions include reliable connectivity, firm development, domain knowledge, and data policies across firms, regions and nation-states. Figure 6 shows the significance of these factors across ID production.

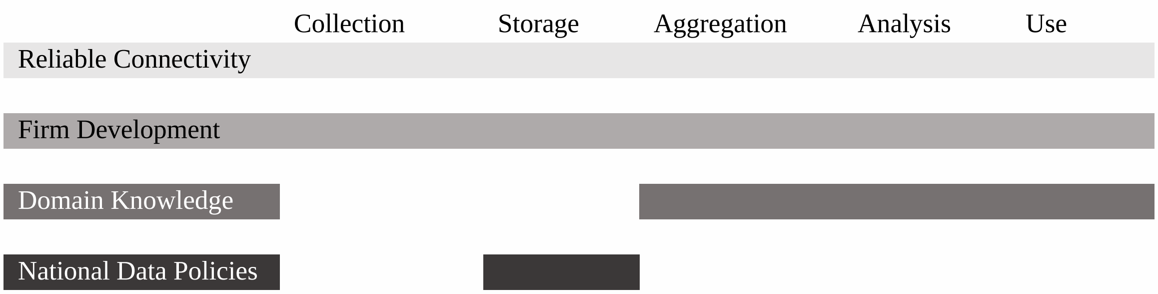


Figure 6. Conditions that influence geography of industrial data production.[31]

As evidenced in the discussion earlier, reliability of connectivity determines how and where data are collected, how frequently the collected data are transferred to a storage system (locally and/or to cloud storage), and who has access to this data. A region with reliable connectivity, including consistent availability of connectivity as well as scalability of bandwidth, as the case may be is likely to support greater mobility of ID across the production chain, and have a competitive advantage in attracting ID-related investment. Similarly, a region with access to both industry domain knowledge and technology and computing skills, and one that supports transfer of this knowledge through relevant industry and research partnerships, is likely to have a competitive advantage in hosting data aggregation and analysis of ID, in supporting the development of relevant, industry-specific use cases of ID. Lastly, national strategy and associated policies of data transfer determine the evolution of this emerging industry.

[30] Developed by research team. The operations that go into these production stages were pieced together from case studies and interviews. Arrows point to the right indicate data flows, and arrows pointing towards left indicate contractual and monetary relationships.
[31] Developed by research team based on analysis of key informant interviews. The figure reveals that the emerging ID industry depends on firm level, regional, and national conditions. For instance, reliable connectivity influences all production functions. Industry domain knowledge is important for aggregating ID with other internal and external information, developing analytical techniques based on sound understanding of how parameters interact with one another, as well as for building use cases specific to the industry sector. National data policies influence how data are transferred across production chain.

Emerging Industrial Data Ecosystem

Delineation of the above factors provides a preliminary picture of the emerging ID ecosystem across the production value chain. As seen in Table 5 below, large industrial firms that manufacture and supply equipment for power generation, tech firms, and automation firms, have a competitive advantage in the emerging industry, owing to the development of internal capacity in this industry over time, and their ability to acquire these capacities via acquisition. Firms such as GE, Siemens, Schneider Electric, as well as Microsoft fall into this category. This classification is much more varied when it comes to

legacy industry sectors. In the context of power generation in vertically integrated utilities in the Southeast, firms with industry domain knowledge are most likely to be develop use cases with support from equipment providers and tech and automation firms. The need for aggregation and analytical services is also giving rise to start-ups within these spaces. See the firm evolution section for more details on these findings.

Table 4. Emerging industrial data ecosystem.[32]

Firm type/ ID production	Collection	Storage	Aggregation	Analysis	Use
Large (incumbent) industrial firms	✓	✓	✓	✓	✓
Large tech and automation firms	✓	✓	✓	✓	✓
Legacy industry sectors (in this case, utilities)					✓
Start-up firms			✓	✓	✓
Industry regulators	✓	✓	✓	✓	✓
Universities and research partnerships				✓	✓
Regional policymakers	✓	✓	✓	✓	✓
National policymakers		✓			

Where the combination of industry domain knowledge and computing capabilities become important - across aggregation, analysis, and use, universities' role in training data scientists, their partnerships to develop high-fidelity analytical models, and their role in incubating new firms through these training and partnership models become important. Similarly, regional and national policymakers also have a role to play, through support for reliable connectivity, domain knowledge transfer, and their data governance policies. See Appendix B for a cumulative summary of these findings, and strategies to fostering regional economic development based on these findings.

Conclusion: Preliminary Strategies and Research Agenda

Corroborating findings from existing literature on cluster studies, the study reveals that the evolution of the emerging ID industry depends on the dynamics and interactions between firms, the institutions that support these firms, and the policy environment. As we have shown, first, these dynamics and interactions exhibit path dependency in terms of an individual firm, the industry sector (here, the energy sector), and the regional economy (e.g. existing manufacturing and knowledge capabilities within the region). Therefore, the emerging ID industry in the Southeast is tied to regional conditions.

[32] Developed by research team.

Consequently, regions with reliable data connectivity, good partnerships among research labs and industry partners, and regional support for research, labor market development, and SMEs can gain a competitive advantage in this emerging industry. However, additional research is required to understand the implications of this emerging industry for the economy in the Southeast. Based on these findings, further scoping of this study should evolve from existing regional industry clusters in the Southeast (see Recommendation #1 [R1] below).

Second, firms in the energy industry have specific domain knowledge needs and operate within a specific regulatory context. This provides some large industrial firms with a greater competitive advantage to evolve within the ID space. Therefore, any policies to support sustainable regional economic development and innovation need to consider pathways for domain knowledge diffusion. Domain knowledge diffusion is important for both spawning new services among existing firms and supporting the development of new firms within the production chain.[33] Technological challenges—such as lack of common standards, data availability, and market value for data across different stakeholders—need to be addressed within the regulatory context and the industry domain in which these concerns exist (see R2).

Domain knowledge includes both the knowledge itself and the process of skilling a workforce with the requisite knowledge. A workforce with knowledge across engineering, computing, and statistics are a crucial component of the emerging ID industry. Subsequent research to explore the changing nature of the workforce within the Southeast is vital to understanding the regional capacity related to ID (see R3). This has implications for how policymakers in the region deal with traffic, immigration, equitable housing, and other issues that go hand-in-hand with a growing population,[34] as well as the relationship between knowledge workers and other economic clusters vital to a healthy economy (The Essential Economy Council, 2013).[35]

Third, the study reveals distributional issues associated with data governance. These issues exist around contractual agreements that define who owns data, who has access to this data, the nature of data-use agreements within firms whose operations are located in different countries, and issues around the influence of national data policies, as well as governance regimes within industries.

In the absence of data governance policies in the emerging ID industry, firms with domain know-how, and greater market share, appear to shape the governance landscape. Therefore, exploring data sharing and use agreements and other contractual agreements would be a useful approach to mapping the emerging ID industry. This will further delineate industry-specific and industry-agnostic concerns and help to sharpen the energy-data nexus in the context of the Southeast (see R4 & R5).

The tension between the goals of different types of firms within power generation, transmission, and distribution, rising concerns about climate change that are empowering sub-national governments to take action on their energy use, also calls for studies on the evolving nature of utility business models in this arena (see R6).

Moreover, the tension between the use of technologies that are intended to overcome geographical barriers and national data policies that seek to strengthen these barriers have implications for the evolution of ID industry. This tension is evident in emerging technological developments such as cloud computing and data localization policies in countries such as China and South Korea.

This mismatch between technology and policies is driven by development priorities on the one hand and security concerns on the other. That is, calls for greater access to data for economic development, and curbing selected types of data sharing due to security concerns, have implications for global governance of data. A future study must address these concerns to understand what a regulatory regime for ID should constitute (see R7).

In conclusion, the emerging ID industry in the context of power generation is the story of evolving technical capabilities, institutional structures, business models, local public policies, and global governance. The interaction between these different elements, and associated externalities of these changes, should not be ignored in any future study.

---

[33] Also see Clark (2013).

[34] Frey (2016) reveals that Florida, North Carolina, Tennessee, and Georgia are home to counties with the some of the highest rates of skilled job growth in the country. In a study of net migration of workers, counties in Tennessee, Florida, and Georgia's Fulton county (in the Atlanta metro region) came out on top during the period 2012-2016.

[35] For example, a 2013 report demonstrated the importance of an economic cluster called the Essential Economy to every county in the State of Georgia. Recent reporting by Clark et al. (2018) highlights the continued role of the Essential Economy in the Southeast and across the entire United States.

Recommendations

1. Scope the study within the context of the Southeast. Consider industry sectors in existing clusters in aviation, energy, automotive manufacturing, OEMs, and logistics (Paul, 2014; Porter, 2005; Waldman and Murray, 2013).
2. Explore domain knowledge providers within the context of existing clusters in the Southeast, to explore industry-specific and industry-agnostic skills and knowledge that would support innovation. That is, how and under what conditions can research institutions, firms, and policymakers support diffusion of domain knowledge in the emerging ID industry?
3. Map existing and new job types and skills that are/will be in demand within the context of the Southeastern region. From a policy perspective, how can existing institutions in the Southeast be leveraged to meet the need for skills in the emerging ID industry?
4. Explore contractual data-use agreements between and within large firms in the Southeast to delineate industry-specific and industry-agnostic data governance challenges. What are the implications for competition and economic development in the emerging ID industry?
5. Are there technology lock-ins that are happening within this space? If yes, are these lock-ins industry specific, or do they span across multiple industry sectors? What are their implications for regional economic development?
6. What are the implications of data-driven changes in policy and society for utility business models?
7. What should a regulatory/global governance regime for ID constitute?

Bibliography

Bernstein, A., Raman, A., 2015. The great decoupling: An interview with Erik Brynjolfsson and Andrew McAfee. Harvard Business Review.

Boria, S., 2015. Developing smart tools for the Airbus Factory of the Future, Industrial internet in action case study. Industrial Internet Consortium.

Bray, E., 2015. Case Study: How Southern Company Predicts Equipment Failure and Reduces Downtime.

Brecht, K., 2015. How U.S. AND EU capital markets are different. Open Markets.

Clark, J., 2013. Working Regions: Reconnecting Innovation and Production in the Knowledge Economy. Routledge, New York.

Clark, J., Lodato, T., Thayer, C., 2018. Bringing the Essential Economy Home: Understanding the Jobs-Housing Mismatch in Sandy Springs.

Coe, N.M., Dicken, P., Hess, M., 2008. Introduction: Global production networks—debates and challenges. J Econ Geogr 8, 267–269. https://doi.org/10.1093/jeg/lbn006

Constine, J., 2012. How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day. TechCrunch.

Cory, N., 2017. Cross-Border data flows: Where are the barriers, and what do they cost? Information Technology and Innovation Foundation, Washington, D.C.

Dicken, P., 2007. Global shift : mapping the changing contours of the world economy, 5th ed. Guilford Press, New York.

Dlodlo, N., 2011. Legal, Privacy, Security, Access and Regulatory Issues in Cloud Computing. Presented at the Proceedings of the 2nd International Conference on Information Management and Evaluation, Ryerson University, Toronto, Canada.

Duke Energy, n.d. Duke Energy renewable services to maintain dte energy's seven wind parks [WWW Document]. Duke Energy. URL https://news.duke-energy.com/releases/duke-energy-renewable-services-to-maintain-dte-energy-s-seven-wind-parkshttp://www.renewableenergyworld.com/articles/2016/11/duke-energy-to-provide-remote-monitoring-control-services-for-offshore-block-island-wind-farm.html (accessed 6.6.17).

Elgendy, N., Elragal, A., 2014. Big Data Analytics: A Literature Review Paper, in: Perner, P. (Ed.), Advances in Data Mining - Theoretical Aspects and Applications, Lecture Notes in Artificial Intelligence. Springer, Germany, pp. 214–228.

EnergyNow, 2015. Big data to the rescue in tracking down electrical theft [WWW Document]. Nextera Energy. URL http://www.nexteraenergy.com/energynow/2015/0715/0715_cover.shtml

Executive Office of the President, 2014. Big Data : Seizing opportunities. Executive Office of the President, Washington D.C.

Ferris, D., Rahim, S., 2016. Tech companies to utilities: Give us your data [WWW Document]. E&E News. URL https://www.eenews.net/stories/1060047592 (accessed 2.10.17).

Frey, W., 2016. U.S. growth rate hits new low as migration to the Sun Belt continues. Brookings Institution

General Electric, 2017. GE Power and Sonelgaz strengthen power sector, mark digital industrial transformation in Algeria through landmark collaboration [WWW Document]. URL http://www.genewsroom.com/press-releases/ge-power-and-sonelgaz-strengthen-power-sector-mark-digital-industrial-transformation

Han, J., Kamber, M., Pei, J., 2012. Data Mining: Concepts and Techniques, 3rd ed, The Morgan Kauffman Series in Data Management Systems. Elsevier Inc., Waltham, MA.

Hashem, A., Yaoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: Review and open research issues. Information Systems 47, 98–115.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed, Statistics. Springer, New York.

Kitchin, R., 2014a. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage.

Kitchin, R., 2014b. Big Data, new epistemologies and paradigm shifts. Big Data & Society 1, 2053951714528481.

Konig, M., 2016. The "emerging issue" of data ownership.

Kusiak, A., 2016. Renewables: Share data on wind energy. Nature News 529, 19. https://doi.org/10.1038/529019a

Loshin, D., 2013. Big Data Analytics : From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. Morgan Kaufmann, Waltham, MA.

Malik, P., 2013. Governing Big Data: Principles and Practices. IBM J. Res. & Dev. 57.

Marr, B., 2015. Big Data In Big Oil: How Shell Uses Analytics To Drive Business Success. Forbes.

Möller, D.P.F., 2016. Guide to Computing Fundamentals in Cyber-Physical Systems. https://doi.org/10.1007/978-3-319-25178-3

O'Donovan, P., Leahy, K., Bruton, K., O'Sullivan, D.T.J., 2015. An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. Journal of Big Data 2, 25. https://doi.org/10.1186/s40537-015-0034-z

Paul, P., 2014. Growing jobs in automotive-related industries: Businesses in Georgia, Alabama, South Carolina, Mississippi and Tennessee will benefit [WWW Document]. Georgia Tech News Center. URL http://www.news.gatech.edu/2014/12/03/growing-jobs-automotive-related-industries (accessed 11.17.15).

Plataine, 2017. Case Study: Harbin Hafei Airbus. Plataine, Waltham, MA.

Porter, M., 2005. South Carolina Competitiveness Initiative: A Strategic Plan for South Carolina. Monitor Company Group.

Porter, M.E., 1998a. Competitive advantage : creating and sustaining superior performance : with a new introduction, 1st Free Press. ed. Free Press, New York.

Porter, M.E., 1998b. On Competition, The Harvard business review book series. Harvard Business School, Boston.

Porter, M.E., 1996. What is strategy? Harvard Business Review 74, 61.

Porter, M.E., 1990. The competitive advantage of nations. Free Press, New York.

Rodman, Meentemeyer, 2006. A geographic analysis of wind turbine placement in Northern California. Energy Policy 34, 2137–2149.

Sassen, S., 2001. The Global City: New York, London, Tokyo. Princeton University Press, Princeton, NJ.

Schwab, K., 2015. The Fourth Industrial Revolution: What it means and how to respond. Foreign Affairs.

Shieber, J., 2017. Looking to sell utility billing data to lenders, Urjanet rakes in $20 million [WWW Document]. TechCrunch. URL https://techcrunch.com/2017/03/02/looking-to-sell-utility-billing-data-to-lenders-urjanet-rakes-in-20-million/ (accessed 9.14.17).

TechNet, SunSpec Alliance, DBL Partners, 2016. Unlocking Grid Data.

The Economist, 2010. Data, data everywhere. The Economist.

The Essential Economy Council, 2013. The Essential Economy Report of Findings: State of Georgia. Atlanta, GA.

Tomer, A., Kneebone, E., Shivaram, R., 2017. Mapping broadband availability and subscription in American neighborhoods. Brookings Institution, Washington, D.C.

Tweed, K., 2010. ABB Acquires Ventyx for Over $1 Billion [WWW Document]. GreenTech Media. URL https://www.greentechmedia.com/articles/read/abb-buys-ventyx-for-over-1-billion (accessed 9.12.17).

University of Cambridge, 2016. Cambridge spin-out secures $3m to improve oil recovery efficiency [WWW Document]. University of Cambridge. URL https://www.enterprise.cam.ac.uk/news/university-of-cambridge-spin-out-secures-3-million-investment-to-improve-the-efficiency-of-oil-recovery/ (accessed 2.25.17).

Verge, J., 2015. 21Vianet to Operate Azure Cloud in China into 2018. Data Center Knowledge.

Waldman, C., Murray, M., 2013. Advanced Manufacturing in the American South : An Economic Analysis Supporting Regional Development. Manufacturers Alliance for Productivity and Innovation.

Wee, D., Kelly, R., Cattel, J., Breunig, M., 2015. Industry 4.0 - how to navigate digitization of the manufacturing sector. McKinsey & Company 1–62. https://doi.org/10.1007/s13398-014-0173-7.2

World Economic Forum, 2016a. Digital Transformation of Industries : Digital Enterprise 45.

World Economic Forum, 2016b. The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution. Growth Strategies 2–3. https://doi.org/10.1177/1946756712473437

Appendices

Appendix A

Presentation of case studies explaining geography of ID production. Developed by research team based on key informant interviews.

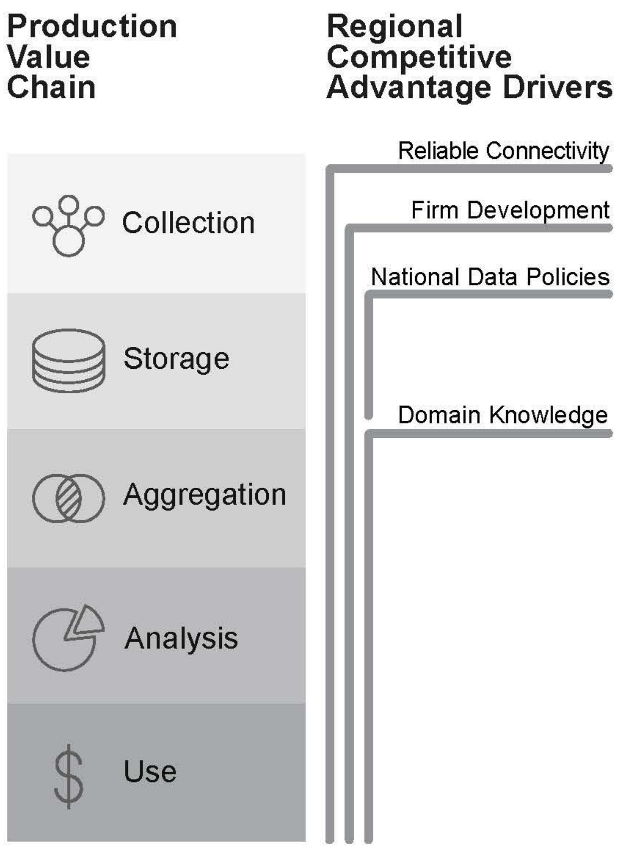
Factors influencing geography of		
Collection	**Storage/Aggregation/Analysis**	**Use**
Data collection undertaken in utility-owned facility due to proprietary nature of data.	While utility possesses domain knowledge, it lacks the resources to integrate equipment data and develop analytical models, etc. Therefore, firm depends on software provided by tech firms and others to aggregate, analyze, and visualize equipment performance.	Data are used internally for ensuring reliability of operations
Collects data from all over its equipment supplied all over the world via maintenance contracts and guarantees. Location of collection depends on reliability of connectivity, latency requirements, for collected data, and national data policies.	The location of facility and for transforming collected data is determined by internal firm decisions undertaken over time. In general, the large incumbent, industrial firm A has leveraged the reliability of connectivity in the Southeast, and partnerships with Georgia Tech to develop analytical models for new data use cases, to train and secure a workforce with both computing and domain	Data are used for providing different types of maintenance services, as well as to develop algorithms and applications that support new use cases. Partnerships with a wide range of stakeholders via its Predix platform supports its business,
Data collection is influenced by utility regulations, need to connect generation assets with aging infrastructure, and reliability of certain equipment and critical infrastructure. Firm’ s visualization software allows utilities to undertake collection at their site.	Domain knowledge needs have led to firm’ s partnership with Georgia tech. The role that Georgia Tech plays in providing resources with data and computing skills, partnership, and incubation space, to work in close collaboration with domain experts within the firm is evident.	Non-utility uses include energy management industrial facilities, data centers, advanced metering behind-the meter, and other critical infrastructure.
Works with data that CW has access to based on forty years of collection, namely, industrial equipment's thermal performance. This is driven by the competition with gas and the industry's "nuclear promise" to reduce cost by 30% over a 5-year term. Location of data collection also depends upon national regulation.	Equipment performance data are aggregated with other data from its customers, to provide focused services. The firm's access to data, subject matter expertise on regulatory issues, licensing, program design have enabled the firm to support their customers in industrial data production. Where data are not allowed to be stored in the cloud, such as in South Korea, a local VPN is utilized, that is not accessible beyond the region. Data storage is not undertaken internally, due to firm's concerns about	Current use cases are largely determined by firm's expertise and data access. For future development, CW acts more as an application developer.

Firm Name – Type				FirmName - type
	Collection	Storage/Aggregation/Analysis		
SME firm A	Firm collects data and builds it on top of customer attribute data that the firm has access to. These include financial metrics, distribution grid assets, and others.	For security purposes, firm deploys its platform at the customer's end, with support from its team of utility experts.		Utilities
Start-up firm, born-global	Data on industrial energy use and other performance parameters are collected using customer credentials, that is, using customer password, or through contract with utility.			Large,Incumbent Industrial Firm A, Diversified TNC
Non-profit organization (proxy for energy management firms)	Does not collect data by itself, but has strategic partnerships with energy management and services firms. Advocates decision makers to improve grid infrastructure, and share different types of grid data with energy management firms.			Large,Incumbent Industrial Firm B, Diversified TNC
SME, (tech) firm B	Data are collected from utility equipment. Depends on customer to provide data for different equipment.	Firm integrates data from different sources, creates data lake, and develops applications for utility customers. This is enabled by the firm's internal capacity development through inclusion of domain experts over time. The firm trains sophomores and graduates from engineering schools in India in order to secure the needed skill development within this space.		Large,Incumbent Industrial Firm C, Diversified TNC

Appendix B

Summary of Findings

Industrial Data & Regional Economic Development: Phase One Findings



Emerging Industrial Data Ecosystem

What are Industrial Data?

Data that are obtained by measuring and assessing the performance of industrial equipment in production and operation.

Sources

Equipment production.
Equipment operation.
Manufacturing and control systems.

Uses

Promote operational reliability.
Develop new firm services & markets.
Promote regional economic development.
Meet sustainability and resilience goals.

Strategies for Regional Economic Development

Ensure reliable data connectivity.

Develop university-industry research partnerships to build high-fidelity, advanced, analytical models.

Support upskilling in the regional labor market.

Investigate evolving data agreements in the emerging ecosystem.

Incubate start-ups.

Lead the development of a global governance framework for industrial data.

Georgia Tech Center for Urban Innovation
Ivan Allen College of Liberal Arts

Georgia Tech Energy Policy and Innovation Center